2000

A Comparative Study of the Performance of Loss Reserving Methods through Simulation

Prakash Narayan
RLI Insurance, pnayaran@rlicorp.com

Thomas Warthen
RLI Insurance Group, twarthen@rlicorp.com

A Comparative Study of the Performance of Loss Reserving Methods through Simulation

Prakash Narayan* and Thomas Warthen†

Abstract‡

Actuaries are often asked to provide a range or confidence level for the loss reserve along with a point estimate. Traditional methods of loss reserving do not provide an estimate of the variance of the estimated reserve, and actuaries use various ad hoc methods to derive a range for the indicated reserve. We use a Monte Carlo simulation method to compare various loss reserve estimation methods, including traditional methods and regression-based methods of loss reserving.

Key words and phrases: *loss development factor, loss triangle, severity, reporting delays, regression, loss ratio*

*Prakash Narayan, Ph.D., A.C.A.S., M.A.A.A., is an actuary at RLI Insurance. He is a graduate of Banaras Hindu University, India and obtained his Ph.D. degree from Monash University, Australia. Dr. Narayan has published several papers in statistical journals. His current research interests include loss reserving methods, property catastrophe loss management, and dynamic financial models of insurance. Dr. Narayan worked for Towers Perrin Reinsurance and Alexander & Alexander as a consultant. His work experience includes loss reserving, development of catastrophe models, and setting retentions and designing reinsurance programs. He is also a member of the American Academy of Actuaries Task Force on Risk-Based Capital.

Dr. Narayan's address is: RLI Insurance, 9025 N Lindbergh Drive, Peoria IL 61615-1431, USA. Internet address: *pnayaran@rlicorp.com*

†Tom Warthen, F.C.A.S., M.A.A.A., is vice president of actuarial services and director of RLI Insurance Group. He earned his B.S. degree in mathematics from the Johns Hopkins University. Prior to joining RLI, Tom was an actuarial consultant at Towers Perrin/Tillinghast, as well as an actuary at Balis & Company, both located in Philadelpia, Pennsylvania.

Mr. Warthen's address is: RLI Insurance, 9025 N Lindbergh Drive, Peoria IL 61615-1431, USA. Internet address: *twarthen@rlicorp.com*

‡The authors would like to thank Janet E. Duncan, F.C.A.S., and an anonymous referee for many helpful suggestions that significantly improved the quality of the presentation.

1 Introduction

Loss reserving, or projecting losses to their ultimate value, is an important actuarial function. The loss development factor (LDF) method attempts to estimate the pattern with which losses for a given cohort of claims change over time. This method produces a point estimate of the required reserve and is the most commonly used actuarial technique for projecting losses to their ultimate value. Actuaries are often asked to provide a range or the variability associated with the point estimate of the loss reserve.

Mack (1994) developed a methodology to estimate the variability of the estimated loss reserves when the LDF method is used. His method may not be appropriate in many situations, however, as the selection of the development factors is often judgmental. Holmberg (1994) has also presented a model by which actuaries can estimate the variability of their loss reserve estimates. Regression modeling of the loss triangle, which can provide both a point estimate and the variability associated with the point estimate, is receiving increasing attention from actuaries. Regression methods provide an estimator of the variance more directly. These methods, however, are rarely used by actuaries because of the methods' complexity. It is desirable to thoroughly test a new methodology before it can be accepted as an appropriate technique and used in practice. Comparisons of forecasting methods based on historical data are not generally considered an objective method for testing forecasting methods. Such studies are likely to be biased by the preference of the investigator.

Alternatively, statistical simulation is a well-accepted technique for comparing various methods of estimation when the properties of the estimators cannot be studied analytically. Stanard (1985) used this technique to compare various traditional methods of loss reserving. We shall apply the same technique to compare the traditional methods with the regression method of loss reserving. Our study uses a variety of methods to simulate the loss triangles.

We have selected the LDF method as one to compare because it is the most commonly used traditional actuarial method. We have included the Bühlmann complementary loss ratio method (which Standard refers to as the additive model), because this method was the best of the tested methods per the Stanard (1985) study. We compare these loss reserve estimation methods and regression methods. The various loss reserve estimation regression models considered in this study differ in the number of the parameters used in modeling the loss triangle.

Our approach is to simulate random loss triangles with a variety of methods and estimate the corresponding loss reserves using the loss development method, Bühlmann complementary loss ratio method, and log-regression models. We assume that the ultimate losses (and hence the reserves) are known with certainty. We compute the deviations of the estimated reserves from the actual reserves derived by various methods. We expect this deviation to be small for a good reserving method. We use several criteria to compare the estimated deviations of actual versus estimated reserves under the various reserving methods.

In the second section the particular methods of simulating random loss triangles are described. We do not claim that these methods capture all the intricacies of the claims process. Our methods also do not generate loss triangles that incorporate the effects of structural changes in the loss process. We also require that incremental losses be positive in our generated triangles. In reality, this constraint may be violated in some actual loss triangles. We believe, however, that our methods generate loss data triangles that are stochastic and do not provide an apparent advantage to any particular method of loss reserve estimation. A particular method of reserve estimation may incorporate some underlying assumptions about the claims process and will obviously provide a better estimate of the loss reserve if those assumptions are valid. In practice it may not be possible to test the assumptions underlying a particular loss reserve estimation method. If a statistical test is applied, it can only detect a gross violation of the assumptions and cannot confirm that those assumptions are true.

Loss development factor methods have an extensive history of use in actuarial practice that preceded the investigation and documentation of the assumptions underlying these methods. Given the current and historical familiarity with loss development factor methods, the assumptions underlying these methods are in some sense secondary to the methods themselves. Given their widespread historical use and technical adequacy as loss development estimation methods, loss development factor methods would be used by actuaries even if no studies about the underlying assumptions were ever published. This is a major consideration, which leads us to use a variety of methods to simulate the random loss triangles.

We are comparing a traditional loss development factor loss reserve estimation method, the Bühlmann complementary loss ratio method, and three fixed regression loss estimation models to estimate the loss reserves. These methods and models are briefly described below. We also discuss the criteria used to compare the results of the simulations. One can definitely define comparison criteria other than those used here. The criteria used are comprehensive, and an estimator performing better in the criteria considered will likely be a good estimator with respect to other reasonable criteria. We have also provided a brief summary of the results of the simulations for the aggregate loss reserves in this section. Appendix A provides the individual accident year results of our computations. We end with several observations of the results and some conclusions based on this simulation study.

2 Simulating Random Loss Triangles

Modeling a claims process to generate the random elements of a loss triangle is complicated. There does not appear to be any study that derives a severity distribution for losses where the individual loss amount may change over time. Stanard (1985) and Pentikäinen and Rantala (1995) describe methods of simulating random loss triangles. Their methods are fundamentally different. The Stanard method is based on a loss severity distribution of individual claim amounts whereas Pentikäinen and Rantala use an aggregate stochastic claim process.

The various methods of loss triangle simulation used here do not satisfy the assumptions underlying the various methods of loss reserving compared. For example, for log-regression modeling, it is assumed that the incremental losses are independent. This assumption is violated by all the methods used for simulating the random loss triangles. Similarly the random loss triangle simulation methods do not satisfy the basic requirement of the LDF method that the future development is determined by the latest available data. One can infer that our study tests the robustness of the various methods of loss reserving against data sets that do not conform to the assumptions underlying the reserve estimation methods.

We have used four different techniques for simulating the loss triangles. The Pentikäinen and Rantala (1995) method is one of them. As we shall see later, the log-regression method of loss reserving requires that the incremental losses be positive. If this is not the case, some subjective judgments need to be made. One way to treat such incidences is to delete such observations from the data set. To be uniform and consistent, we have selected loss triangle simulation methods that will generate positive incremental losses. Stanard's method does not satisfy this requirement and is not used.

For all the methods in this study, 11 accident years are considered. It is further assumed that the losses completely mature at the 11th year of development, i.e., the first accident year is at the ultimate loss level and no further development is expected. Because we require complete knowledge of the ultimate losses for a proper comparison of the reserve estimation results of the different estimation methods, we generate a complete history for each accident year. In estimating the reserves, only the top half of the loss triangle is available to the actuary as data. The top half is used to estimate the lower half of the triangle, particularly the last (right) column, which represents the projection of ultimate losses. For $i, j = 1, 2, \ldots$, let $S_{i,j}$ denote the incremental losses for the accident year i at the end of the calendar year $i + j - 1$, and let $L_{i,j}$ denote the cumulative losses for the accident year i at the end of the calendar year $i + j - 1$, i.e.,

$$L_{i,j} = \sum_{k=1}^{j} S_{i,k}.$$

The ultimate loss for accident year i, L_i, is given by

$$L_i = \lim_{j \to \infty} L_{i,j}.$$

For simplicity the losses are assumed to be fully developed after 11 years, i.e., $L_i = L_{i,11}$. In addition, we consider only 11 accident years.

2.1 Random Reporting Factor

The steps of this random loss triangle generation method for accident year i ($i = 1, \ldots, 11$) are:

Step 1: Generate N_i, the number of losses for accident year i, as a Poisson random variable with mean 100.

Step 2: Generate N_i claim amount variables $\{C_{i,1}, C_{i,2}, \ldots, C_{i,N_i}\}$ where each $C_{i,k}$ is log-normally distributed with parameters $\mu = 7.3659$ and $\sigma = 1.517427$.[1] These parameters correspond to a loss severity mean of 5000 and a coefficient of variation of 3. The ultimate losses for accident year i is

$$L_i = L_{i,11} = 1.06^{(i-1)} \sum_{k=1}^{N_i} C_{i,k}.$$

[1] A random variable X is said to be log-normally distributed with parameters μ and σ if $\ln X$ is normally distributed with mean μ and variance σ^2.

Step 3: Generate ten random numbers $U_{i,j}$, for $j = 1, \ldots, 10$, that are uniform on $(0, 1)$.

Step 4: For $j = 1, \ldots, 10$ compute

$$T_{i,j} = \frac{1}{10} + \frac{1}{2}U_{i,j} + \frac{1}{2}\ln(j) \tag{1}$$

and

$$X_{i,j} = \sum_{k=1}^{j} T_{i,k}. \tag{2}$$

Step 5: The simulated cumulative loss for accident year 1 at lag (delay) j, $L_{1,j}$ is given by

$$L_{i,j} = L_{i,11}(1 - e^{-X_{i,j}}).$$

Note that the accident year losses are inflated by 6 percent per year.

Though this method may look like a development factor model, it does not strictly satisfy the assumptions of the loss development factor model. The ratio of the expected losses $E[L_{i,j+1}]/E[L_{i,j}]$ is a constant, not the conditional expectation. It also does not satisfy the assumption of independence of incremental losses underlying the log regression models of loss reserves.

2.2 Random Backward Development Factor

This method is similar to method 1 except that the factors are computed in reverse order. The steps of the method for accident year i are:

Step 1: Generate N_i, the number of losses for accident year i, as a Poisson random variable with mean 100.

Step 2: Generate N_i claim amount variables $\{C_{i,1}, C_{i,2}, \ldots, C_{i,N_i}\}$ where each $C_{i,k}$ is log-normally distributed with parameters $\mu = 7.3659$ and $\sigma = 1.517427$. The ultimate loss for accident year i is

$$L_i = L_{i,11} = 1.06^{(i-1)} \sum_{k=1}^{N_i} C_{i,k}.$$

Step 3: For $j = 1, 2, \ldots, 10$, generate the log-normal variates $Y_{i,11-j}$ with parameters $\mu_{i,j} = a_j$ and $\sigma_{i,j} = b_j$ where

$$a_j = \frac{(j + (j-1)^2)}{100}$$

and

$$b_j = \frac{(j + (j-1)^2)}{500}.$$

Note that $Y_{i,j}$ is a randomly generated development factor for the development period j to $j+1$.

Step 4: Losses reported at the end of year 10 for the accident year i, $L_{i,10}$, are $L_i/Y_{i,10}$. The reported losses at earlier valuation dates are computed by dividing by $Y_{i,j}$ successively, i.e.,

$$L_{i,j} = \frac{L_{i,j+1}}{Y_{i,j}}, \quad j = 10, 9, \ldots, 2, 1.$$

The a_j and b_j parameters are selected so that $\Pr[Y_{i,j} > 1] = 1 - \epsilon$ for very small ϵ.

2.3 Individual Losses with Changing Severity

This method is based on the ideas of Stanard (1985) and Bühlmann, Schnieper, and Straub (1980). As in Stanard, we assume an exponential delay in reporting and settlement with the added assumption that the severity distribution varies with delay. The claim amounts are assumed to follow a Pareto distribution with parameters λ and θ.[2]

[2] A random variable X is said to have a Pareto distribution with parameters λ and θ if

$$\Pr[X \le x] = 1 - (1 + \frac{x}{\lambda})^{-\theta} \quad x > 0.$$

As each individual claim develops, the percentile level of the individual loss is assumed to remain constant over time but the parameters λ and θ are assumed to change until the claim is settled. In other words, if the kth claim in accident year i is initially of size $C_{i,k}$, the percentile level of the claim is $U_{i,k}$ where

$$U_{i,k} = 1 - \left(1 + \frac{C_{i,k}}{\lambda}\right)^{-\theta}.$$

For $k = 1, 2, \ldots, N_i$, the kth claim in accident year i, $C_{i,k}$, is assumed to have three random characteristics measured from the beginning of the accident year: the date of occurrence, $X_{i,k,1}$, which is a uniform variate on $(0, 1)$; the reporting delay, $X_{i,k,2}$, which is exponentially distributed with mean 2; and the settlement delay, $X_{i,k,3}$, which is exponentially distributed with mean 2. As we require that the ultimate values of a claim be known within 11 calendar years after it occurred, we truncate both $X_{i,k,1} + X_{i,k,2}$ and $X_{i,k,1} + X_{i,k,2} + X_{i,k,3}$ at 11 if they exceed 11. This provides loss amounts for each claim for delays for $j = 1, 2, \ldots, 11$. Specifically, let $r_{i,k}$ and $R_{i,k}$ be nonnegative integers such that

$$r_{i,k} = \min\{\lfloor (X_{i,k,1} + X_{i,k,2}) \rfloor,\ 11\} \tag{3}$$

$$R_{i,k} = \min\{\lfloor (X_{i,k,1} + X_{i,k,2} + X_{i,k,2}) \rfloor,\ 11\} \tag{4}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. It follows that the kth claim in accident year i is reported in calendar year $i + r_{i,k}$ and is settled in calendar year $i + R_{i,k}$. The estimated loss after delay j is $\hat{C}_{i,k,j}$, which is defined as:

$$\hat{C}_{i,k,j} = \begin{cases} 0 & \text{if } j = 1, 2, \ldots, r_{i,k}; \\ \lambda(j)\left(\dfrac{1}{(1 - U_{i,k})^{1/\theta(j)}} - 1\right) & \text{if } j = r_{i,k} + 1, \ldots, R_{i,k}; \\ \lambda(j)\left(\dfrac{1}{(1 - U_{i,k})^{1/\theta(R_{i,k})}} - 1\right) & \text{if } j = R_{i,k} + 1, \ldots, 11; \end{cases} \tag{5}$$

where $\lambda = \lambda_1$, $\theta = \theta_1$

$$\lambda(j) = 50(20 + j - 1)(1.06)^{j-1}$$

and

$$\theta(j) = (50 - (j - 1))/20.$$

Note that the middle expression for $\hat{C}_{i,k,j}$ in equation (5) can be written as

$$\left(\frac{1}{(1-U_{i,k})^{1/\theta(j)}}-1\right)=\left(\left(1+\frac{C_{i,k}}{\lambda}\right)^{\theta/\theta(j)}-1\right)$$

and that if a claim is settled at delay j then $\hat{C}_{i,j,k}$ remains constant at later valuations.

The estimated claim amount $\hat{C}_{i,k,j}$ increases over time (as j increases). In actual practice, however, the estimated claim amount may decrease from an earlier to a later valuation for some claims. The procedure used here will always increase the severity of the loss from one valuation to the next. This is done to force the incremental losses to be positive.

The steps of the method for accident year i are:

Step 1: Generate N_i, the number of losses for accident year i, as a Poisson random variable with mean 100.

Step 2: Generate N_i claim amount variables $\{C_{i,1}, C_{i,2}, \ldots, C_{i,N_i}\}$ where each $C_{i,k}$ is a Pareto distribution with parameters $\lambda = 1000$ and $\theta = 2.5$. The corresponding $\{U_{i,1}, U_{i,2}, \ldots, U_{i,N_i}\}$ are also determined.

Step 3: For the $k = 1, 2, \ldots N_i$, generate uniform $(0,1)$ variates $X_{i,k,1}$ for the occurrence date and exponential variates $X_{i,k,2}$ and $X_{i,k,3}$ with mean 2 and 5 respectively for the reporting delay and the settlement delay. The quantities $r_{i,k}$ and $R_{i,k}$ are calculated according to equations (3) and (4).

Step 4: Calculate the $\hat{C}_{i,k,j}$s for $j = 1, 2, \ldots 11$. Note that the ultimate loss for accident year i is

$$L_i = L_{i,11} = 1.06^{(i-1)} \sum_{k=1}^{N_i} \hat{C}_{i,k,11}. \tag{6}$$

2.4 Pentikäinen-Rantala Method

This method is based on the procedure described by Pentikäinen and Rantala (1995). Our implementation differs slightly from theirs. We shall describe the computational steps of this method briefly; the reader is encouraged to review the original Pentikäinen-Rantala paper for a complete explanation of their method. The computational steps of this method are:

Step 1: We assume a reporting pattern for a cohort of aggregate losses. This pattern is assumed not to change over time and includes pure IBNR. Specifically, let $X(j)$ denote the proportion of the losses in accident year i reported in calendar year $i+j-1$, $j = 1, 2, \ldots, 11$. The pattern used is $X(1) = 0.220$, $X(2) = 0.180$, $X(3) = 0.150$, $X(4) = 0.120$, $X(5) = 0.100$, $X(6) = 0.080$, $X(7) = 0.060$, $X(8) = 0.040$, $X(9) = 0.027$, $X(10) = 0.016$, $X(11) = 0.007$.

Step 2: Claims for the accident year i reported at delay j are given by

$$S_{i,j} = K \times X(j) \times XP(i) \times q(i,j) \times \mathrm{INF}(i+j-1) \quad (7)$$

where K is constant parameter related to the total losses for accident year 1;

$$XP(i) = ((1.01)(1.06))^{i-1} \quad \text{Exposure and inflation growth;}$$
$$q(i,j) = 0.4 + 0.6q(i,j-1) + \epsilon_{i,j}$$

where $q(i,0) = 1$ and $\epsilon_{i,j} \sim N(0, 0.05)$

$$\mathrm{INF}(t) = \prod_{k=1}^{t} (1 + \delta(k));$$
$$\delta(k+1) = \max(0.06 + 0.7(\delta(k) - 0.06) + \omega_k)$$

and $\delta(1) = 0.06$ and $\omega_k \sim N(0, 0.015)$.

This method is based on randomizing the aggregate losses of all the claims for an accident year. Claim reporting and inflation are modeled by autoregressive processes. We further restrict the inflation rate to a minimum of 3 percent. This method also has an exposure growth of 1 percent.

We note that in the simulation of random loss triangles by the methods of Sections 2.1, 2.2, and 2.3, individual claim severity is unlimited. In practice individual losses will have an upper limit in most cases. Occurrence of an individual large loss in the simulation process may cause an individual accident year loss to be out of line with other accident year losses in an individual loss triangle.

It is worth stating that the computations for the simulations were performed in Excel. We have, however, implemented our own module to generate the uniform random variate.

3 Methods of Loss Reserving

One can see that each of the four methods of generating loss triangles in Section 2 has several parameters. As these parameters are changed, the simulated triangles may exhibit significantly different development patterns. A particular method of loss reserving, considered best with a selected set of loss triangle generation parameters, need not be better for any other set of loss triangle generation parameters. The simulation conducted here emphasizes a variety of methods of loss triangle generation rather than the sensitivity of the loss triangle generation methods over a range of possible parameters.

Let us assume that there is no further claim development beyond year n or, equivalently, that $L_{i,n}$ is the ultimate loss value for the accident year i. (Recall in Section 2 that $n = 11$.) Further assume that all $S_{i,j}$ are positive and let

$$Z_{i,j} = \ln(S_{i,j}). \tag{8}$$

To simplify the later exposition of our estimation process, let us further assume that the accident year loss inflation rate is 6 percent and there is no exposure growth except for the Pentikäinen-Rantala method in which constant exposure growth of 1 percent is assumed. Our problem is to estimate $L_{i,j}$ for $i = 1, 2, \ldots, n$ and $j = n+2-i, n+3-i, \ldots, n$ given that $L_{i,j}$ is known for $i = 1, 2, \ldots, n$ and $j = 1, \ldots, n+1-i$.

Two traditional methods of loss reserving and three regression models are used. The two traditional methods are the loss development factor method and the Bühlmann complementary loss ratio method. The loss development factor method is the most commonly used actuarial technique. The Bühlmann loss ratio method was chosen for this analysis because this method outperforms other actuarial methods in the simulation study by Stanard (1985).

The three regression models we have selected for comparison are similar; the differences among them lie in the number of parameters fitted. These methods are described next.

Loss Development We compute:

$$f_{i,j} = \frac{L_{i,j+1}}{L_{i,j}}$$

$$f_j = \frac{1}{n-j} \sum_{i=1}^{n-j} f_{i,j}$$

$$u_k = \prod_{j=k}^{n-1} f_j$$

and the estimated $\bar{L}_{i,n}$ is given by

$$\bar{L}_{i,n} = L_{i,n-i+1} u_{n-i+1}.$$

Bühlmann Complementary Loss Ratio Method This method of loss reserving has not been commonly applied in North America and is suitable for application to paid loss data. It is based on the presumption that the proportion of losses paid at a particular delay remains constant over time. This proportion is estimated from the historical loss experience and is used to forecast the future. We compute:

$$\bar{M}_j = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} S_{i,j}(1+r)^{n-i} \quad \text{for } j = 2, 3, \dots, n \text{ and}$$

$$S_{i,j} = \bar{M}_j(1+r)^{i-n} \quad \text{for } j = n+2-i, \dots, n \text{ and } i = 2, 3, \dots, n$$

where r is rate of inflation for losses and is assumed to be 6 percent in our simulation.

Regression Models Our discussion of the regression models considered in our analysis is brief. These models are discussed in greater detail by Zehnwirth (1994) and Verrall (1994) among others. We have used an unbiased estimator for the loss reserves as recommended by Verrall (1994) rather than Bayes or maximum likelihood estimates (MLE). In these models the incremental losses are assumed to follow some stochastic distribution. Usually some transformation is applied to the incremental losses before the model parameters are estimated. Although various transformations have been investigated, the logarithmic transformation is most commonly used. Let us describe the methodology briefly with the log transformation for completeness. Readers not familiar with the methodology are encouraged to review the papers by Verrall (1994) and Zehnwirth (1994).

Recall equation (8). We assume that

$$Z_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j}$$

where μ, α_i, and β_j are the constant parameters of the model and the $\epsilon_{i,j}$ are error terms that are assumed independent identically distributed normal variates with mean 0 and variance σ^2 and are the error terms or the random noise. We make the usual assumption that $\alpha_1 = 0$ and $\beta_1 = 0$ to make the model of full rank. The parameters of the model are estimated by the least squares method. Under the assumption of the normality of the error terms, the estimates are also MLEs. We use the unbiased estimate for the forecasting and require that the errors are independent and normally distributed.

The three regression models investigated in this paper are:

Model 1: α_i and β_j are all different for $i, j = 2, 3, \ldots, n$.

Model 2: $\alpha_i = (i - 1)\alpha$ for $i = 2, \ldots, 11$ and β_j are different for $j = 2, 3, \ldots, n$.

Model 3: $\alpha_i = (i-1)\alpha$ and $\beta_j = (j-1)\beta + \gamma \ln(j)$ for $i, j = 2, 3, \ldots, n$.

In the actual application of regression models, one will select the model that provides the best fit to the data based on the evaluation of the residuals and other statistics of the fitted models. Such an approach is not feasible in simulation. Zehnwirth (1994) emphasizes parsimony when applying the regression models for forecasting. The number of parameters used in the three regression models is 21, 12, and 4, respectively. The difference among the three models lies in the number of parameters used to fit the data. As defined, regression model 1 has too many parameters and model 3 too few to capture the essence of a random loss triangle. Model 2 and model 3 assume some underlying relationships among the model 1 parameters. In selecting these regression models, our purpose is not to compare these models with each other, but to see the effect of using fewer parameters in regression modeling.

The parameters are estimated by the least squares method and used to forecast ultimate losses. We refer the reader to Verrall (1994), who provides a complete description of the estimation method and an unbiased estimator of the lower triangle for model 1. The other regression models require revisions to the design matrix and modification of the appropriate equations from those described in Verrall (1994).

4 Comparison of Procedures

We have generated 5000 hypothetical loss triangles for each of the simulation methods described earlier. For each of the 5000 sets of hypothetical data, the reserves are estimated by the loss development method, Bühlmann complementary loss ratio method, and the regression loss reserve estimation methods. The deviations between the loss reserve estimates and the actual reserves are computed.

An important property of a good estimator is that it is unbiased. Stanard (1985) used this criterion for comparing various loss reserve estimators. If an estimator is unbiased, the average deviation of estimated versus actual reserves over many simulations will be negligible.

Between two unbiased estimators, statisticians prefer the estimator with the smaller variance. Between biased estimators, the estimator with the minimum mean square error is preferred. In our context, this means that the average squared deviations between the estimated and actual reserves should be small. This is an important criterion for a reserve estimation method in the insurance context.

The reserves are an important component of the insurer's financial reporting. A reserving method that provides estimates with small biases, but for which the individual simulation (data set) estimates vary a lot from the actual reserves, may not be an appropriate reserve estimation method. One will prefer the reserve estimates to be closer to the true value. We use root mean square error (RMSE) and the average absolute deviation of the estimated versus the actual reserve to test the closeness of the reserve estimators to the actual reserve values. We also compute the average percentage error. A reserve estimation method that generates a smaller percentage error in the estimate is better. Another criterion used to compare the various loss reserving methods is to compute the correlation between the actual reserves and the estimated reserves. One would expect a high correlation for a good reserving method.

We compare the reserve estimates for each of the loss reserve estimation methods, for each of the random loss triangle simulation methods. Our comments follow:

Random Reporting Factor: The Bühlmann complementary loss ratio method is the best loss reserve estimation method for the random reporting factor method of random loss triangle simulation. The regression models of loss reserve estimation perform better than the loss development factor method. The correlation for all the reserving methods is low and surprisingly is smallest for the Bühlmann complementary loss ratio method. Regression model 2 performs slightly better than regression model 1. The main difference between these models is that regression model 2 estimates accident year inflation and allows one parameter for that model component, whereas regression model 1 allows an inflation parameter for each accident year. Our results indicate that parsimony in the regression model is important and that overparametrization may provide inferior results.

Random Backward Development Factor: Regression model 3 appears to be the best method for this loss simulation method based on aggregate combined accident years' forecast. The Bühlmann complementary loss ratio method is superior in the individual accident year forecasts. Regression model 3 does not capture the payout pattern correctly. The other regression models perform better than the loss development factor method. The Bühlmann method again shows poor correlation with the actual reserves, while the other methods show a reasonable correlation level. We conclude that the Bühlmann method and regression model 2 perform better for this method of random loss triangle simulation than the other tested methods.

Individual Losses with Changing Severity: The loss development method performs well for this loss simulation method. Regression model 2 appears to be better overall. Regression model 3 performs poorly, perhaps because of an insufficient number of model parameters.

Pentikäinen and Rantala Method: Regression models 1 and 2 outperform the other methods. The loss development factor method performs better than the Bühlmann complementary loss ratio method and regression model 3.

Table 1 summarizes our results for each of the four methods of random simulation of the hypothetical loss triangles. Tables 2 through 5 provide similar statistics for individual accident years. These tables show that one of the three regression models considered in this analysis is generally better than the LDF method.

Table 1
Summary of Results the Four Methods
Of Random Simulation of Hypothetical Loss Triangles

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
Five Thousand Iterations Under Method 1					
Actual Total Reserve: Average = 1,108,298, Std. Dev. = 244,287					
Bias	151,681	5,222	36,486	31,240	51,367
RMSE	466,055	266,874	395,819	328,870	341,537
AAD	364,628	204,674	314,829	254,069	263,444
APE	16.84%	4.84%	6.22%	6.75%	8.69%
CORR	0.25	0.09	0.25	0.15	0.14
Five Thousand Iterations Under Method 2					
Actual Total Reserve: Average = 3,665,734, Std. Dev. = 485,206					
Bias	157,684	(8,088)	55,356	15,393	3,125
RMSE	512,092	639,187	481,727	542,257	519,705
AAD	391,022	485,769	373,282	420,438	403,056
APE	4.38%	1.23%	1.58%	0.79%	0.47%
CORR	0.70	0.11	0.70	0.57	0.58
Five Thousand Iterations Under Method 3					
Actual Total Reserve: Average = 1,634,559, Std. Dev. = 252,631					
Bias	30,566	(83,039)	(144,192)	(52,327)	(176,089)
RMSE	413,137	441,109	375,367	299,099	340,506
AAD	356,932	347,340	314,629	259,057	280,243
APE	1.39%	-4.36%	-9.49%	-3.31%	-9.52%
CORR	0.62	0.39	0.68	0.66	0.32
Five Thousand Iterations Under Method 4					
Actual Total Reserve: Average = 3,183,654, Std. Dev. = 330,776					
Bias	10,106	(21,441)	5,326	4,789	34,136
RMSE	186,688	186,916	183,351	195,148	201,012
AAD	147,536	147,830	145,029	153,675	157,283
APE	0.23%	-0.24%	0.07%	0.06%	0.98%
CORR	0.89	0.84	0.89	0.88	0.88

Notes: Loss Dev. = Loss Development; Std. Dev. = Standard Deviation; RMSE = Root Mean Square Error; AAD = Average Absolute Deviation; APE = Average Percentage Error; and CORR = Correlation between the Actual Reserves and the Estimated Reserves.

Table 2
Random Reporting Factor

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Bias				
2	0	0	(0)	0	(0)
3	1	1	(1)	(1)	3
4	3	3	(5)	(4)	7
5	(1)	3	(26)	(23)	(17)
6	21	12	(68)	(62)	(144)
7	60	(10)	(211)	(158)	(607)
8	371	(60)	(383)	(105)	(1,228)
9	1,365	(121)	(844)	1,135	665
10	9,755	1,141	985	7,587	14,377
11	140,106	4,253	37,041	22,871	38,312
Total	151,681	5,222	36,486	31,240	51,367
AY	RMSE				
2	7	8	4	8	5
3	25	31	17	29	25
4	101	122	72	116	107
5	388	478	291	451	425
6	1,371	1,772	1,066	1,662	1,583
7	4,478	6,041	3,722	5,660	5,390
8	13,713	18,441	12,389	17,657	16,815
9	37,972	51,449	37,716	51,583	50,046
10	110,938	117,820	118,183	125,616	126,393
11	441,193	218,265	368,764	258,292	265,187
Total	466,055	266,874	395,819	328,870	341,537

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 2 (continued)
Random Reporting Factor

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Average Absolute Deviations				
2	5	6	3	6	4
3	19	23	12	21	18
4	74	90	51	85	79
5	288	356	209	332	312
6	1,018	1,295	772	1,198	1,119
7	3,373	4,430	2,750	4,150	3,876
8	10,440	13,784	9,320	13,206	12,359
9	29,290	37,826	28,973	38,281	37,003
10	86,296	88,748	91,752	96,232	97,421
11	346,382	166,051	296,479	197,807	203,655
Total	364,628	204,674	314,829	254,069	263,444
AY	Average Percentage Errors				
2	25.67%	36.00%	9.95%	35.38%	19.36%
3	22.20%	33.13%	9.27%	29.45%	35.19%
4	20.02%	29.68%	8.80%	25.10%	29.48%
5	16.10%	26.60%	7.65%	21.81%	21.16%
6	14.12%	23.31%	7.46%	19.31%	15.14%
7	11.63%	20.28%	6.92%	17.33%	11.96%
8	10.17%	17.67%	7.00%	16.12%	12.08%
9	8.67%	15.01%	6.57%	15.10%	14.19%
10	9.89%	13.16%	7.30%	14.80%	17.14%
11	30.23%	10.57%	12.90%	13.69%	16.36%
Total	16.84%	4.84%	6.22%	6.75%	8.69%

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 3
Random Backward Development Factor

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Bias				
2	2	11	(22)	(22)	16,639
3	119	208	69	40	32,153
4	222	360	156	(54)	38,008
5	694	(470)	534	(1,165)	27,398
6	1,321	(904)	664	(1,643)	2,799
7	3,705	(643)	1,683	(678)	(26,084)
8	8,940	(3,630)	3,237	(2,019)	(47,440)
9	23,597	1,501	12,390	5,744	(37,721)
10	42,679	(1,854)	16,585	5,805	(15,255)
11	76,406	(2,667)	20,062	9,384	12,628
Total	157,684	(8,088)	55,356	15,393	3,125
AY	RMSE				
2	1,564	2,837	1,591	2,786	17,107
3	4,614	9,596	4,791	9,176	33,712
4	11,084	24,442	11,681	22,880	44,274
5	22,705	50,969	24,318	46,958	52,965
6	40,045	87,472	43,509	78,937	77,117
7	64,981	133,834	70,337	121,306	120,900
8	105,093	188,472	111,145	171,608	174,206
9	154,998	222,490	157,449	210,056	208,665
10	231,105	269,200	222,925	271,969	268,947
11	320,345	287,746	293,720	318,988	317,808
Total	512,092	639,187	481,727	542,257	519,705

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 3 (continued)
Random Backward Development Factor

	Forecast Method				
	Loss Dev.	Bühlmann	Regression		
	Method	Loss Ratio	Model 1	Model 2	Model 3
AY	Average Absolute Deviations				
2	1,187	2,082	1,206	2,056	16,639
3	3,553	7,119	3,681	6,848	32,192
4	8,525	18,316	8,996	17,259	39,640
5	17,267	37,913	18,525	34,859	43,380
6	30,623	63,847	33,027	57,937	57,143
7	49,699	99,623	53,807	89,929	86,860
8	78,098	133,685	82,422	122,794	119,958
9	115,161	163,563	118,176	156,441	151,540
10	167,451	193,324	163,199	200,545	196,412
11	233,783	211,057	219,487	240,241	239,523
Total	391,022	485,769	373,282	420,438	403,056
AY	Average Percentage Errors				
2	4.73%	13.54%	4.23%	13.27%	360.75%
3	3.43%	12.41%	3.18%	11.90%	171.65%
4	2.61%	11.08%	2.51%	10.42%	77.29%
5	2.30%	9.75%	2.21%	8.90%	31.55%
6	2.01%	8.80%	1.74%	7.86%	9.62%
7	2.25%	9.06%	1.74%	8.10%	0.62%
8	2.65%	8.33%	1.55%	7.37%	-2.07%
9	4.37%	8.81%	2.64%	8.09%	0.93%
10	6.14%	8.52%	2.75%	8.36%	5.41%
11	9.19%	8.02%	2.65%	8.63%	9.01%
Total	4.38%	1.23%	1.58%	0.79%	0.47%

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 4
Individual Losses with Changing Severity

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Bias				
2	1,048	3,668	788	3,751	(9,394)
3	2,867	(5,109)	508	(5,560)	(19,900)
4	(25,022)	(28,645)	(27,849)	(28,329)	(43,090)
5	36,381	(18,287)	17,366	(17,199)	(31,036)
6	29,767	40,903	24,706	39,074	23,710
7	(64,522)	(77,462)	(72,000)	(80,096)	(93,727)
8	40,487	6,886	16,237	6,239	(2,950)
9	3,817	9,224	(7,890)	13,152	8,934
10	8,224	41,085	(9,994)	52,394	46,612
11	(2,480)	(55,304)	(86,063)	(35,753)	(55,246)
Total	30,566	(83,039)	(144,192)	(52,327)	(176,089)
AY	RMSE				
2	31,330	36,022	30,391	35,694	14,678
3	52,869	43,391	51,745	43,108	38,400
4	115,037	126,487	114,408	125,039	118,422
5	171,424	78,184	130,753	78,614	75,765
6	61,614	69,934	58,478	64,412	40,414
7	222,605	230,023	221,536	232,337	262,153
8	110,891	107,524	86,577	95,115	80,851
9	117,997	55,953	119,472	77,183	74,298
10	94,518	97,249	96,401	90,863	86,043
11	259,037	222,586	260,944	173,258	169,985
Total	413,137	441,109	375,367	299,099	340,506

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 4 (continued)
Individual Losses with Changing Severity

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Average Absolute Deviations				
2	19,830	24,090	18,965	24,010	12,509
3	37,669	31,125	36,969	31,640	24,974
4	72,059	77,787	70,388	76,060	62,687
5	105,641	59,338	88,162	61,260	54,471
6	35,254	45,292	33,024	44,186	31,115
7	111,397	120,745	106,008	113,100	119,757
8	94,362	87,744	76,900	78,766	66,943
9	90,948	48,065	88,693	66,927	63,907
10	81,845	75,610	77,164	75,122	66,540
11	202,172	161,009	209,386	135,948	127,446
Total	356,932	347,340	314,629	259,057	280,243
AY	Average Percentage Errors				
2	47.87%	94.00%	44.07%	95.99%	-17.10%
3	49.86%	20.49%	42.22%	20.10%	-13.53%
4	57.18%	63.35%	48.78%	62.45%	16.07%
5	119.19%	12.26%	80.54%	16.42%	0.49%
6	56.20%	77.73%	48.35%	74.57%	48.74%
7	5.03%	-0.81%	0.06%	-3.61%	-7.82%
8	42.12%	29.37%	27.35%	25.73%	17.56%
9	15.32%	8.67%	9.34%	14.34%	13.17%
10	6.43%	21.75%	0.22%	25.76%	23.66%
11	20.04%	3.88%	-2.87%	5.32%	-1.39%
Total	1.39%	-4.36%	-9.49%	-3.31%	-9.52%

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 5
Pentikäinen and Rantala Method

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Bias				
2	2	1	(2)	11	6,037
3	26	(44)	12	11	9,160
4	114	(52)	78	118	9,389
5	283	(217)	213	190	7,237
6	313	(680)	197	162	(755)
7	456	(1,090)	222	473	(9,710)
8	1,225	(2,113)	855	549	(14,271)
9	1,837	(3,756)	1,257	497	(7,450)
10	3,050	(5,083)	1,984	1,456	7,505
11	2,801	(8,407)	509	1,322	26,995
Total	10,106	(21,441)	5,326	4,789	34,136
AY	RMSE				
2	672	608	620	607	6,101
3	1,897	1,724	1,759	1,733	9,399
4	4,070	3,650	3,822	3,687	10,222
5	7,375	6,387	6,997	6,478	9,896
6	12,478	10,855	11,986	10,987	10,942
7	19,710	17,324	19,117	17,606	19,634
8	29,727	26,563	29,027	27,007	29,730
9	40,540	37,157	39,974	37,662	37,813
10	54,967	51,969	54,481	53,464	54,281
11	74,329	71,882	73,934	76,120	82,500
Total	186,688	186,916	183,351	195,148	201,012

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

Table 5 (continued)
Pentikäinen and Rantala Method

	Forecast Method				
	Loss Dev. Method	Bühlmann Loss Ratio	Regression Model 1	Regression Model 2	Regression Model 3
AY	Average Absolute Deviations				
2	536	482	493	480	6,037
3	1,499	1,363	1,388	1,368	9,160
4	3,229	2,912	3,029	2,924	9,401
5	5,839	5,070	5,547	5,105	8,119
6	9,887	8,663	9,495	8,759	8,728
7	15,628	13,742	15,153	13,986	15,864
8	23,414	21,143	22,868	21,350	23,951
9	31,748	29,300	31,319	29,800	30,088
10	43,420	41,011	43,107	41,990	42,469
11	58,767	56,865	58,504	60,077	64,294
Total	147,536	147,830	145,029	153,675	157,283
AY	Average Percentage Errors				
2	0.42%	0.51%	0.32%	0.56%	89.77%
3	0.38%	0.25%	0.29%	0.34%	40.30%
4	0.44%	0.35%	0.35%	0.48%	18.60%
5	0.46%	0.23%	0.37%	0.38%	7.81%
6	0.30%	0.06%	0.21%	0.25%	-0.35%
7	0.25%	0.08%	0.15%	0.30%	-3.72%
8	0.37%	0.00%	0.26%	0.25%	-3.71%
9	0.34%	-0.13%	0.23%	0.15%	-1.36%
10	0.40%	-0.09%	0.25%	0.22%	1.06%
11	0.26%	-0.22%	0.02%	0.16%	2.82%
Total	0.23%	-0.24%	0.07%	0.06%	0.98%

Notes: Loss Dev. = Loss Development; AY = Accident Year;
RMSE = Root Mean Square Error.

The Bühlmann method is slightly better in some cases, but we assumed that the inflation rate is known for the Bühlmann method. We are therefore using additional information for this method and are obtaining slightly better answers. Such information will ordinarily not be available in actual practice. Actual loss data will be tainted by both exposure changes and the inflationary loss cost changes that will vary over time. For most of the methods of random loss generation the effect of inflation has a minimal impact on the ultimate answer derived by traditional methods. Inflation affects the weighting given to individual accident years in the total reserve. For regression models, inflation will affect the forecast in a more complicated fashion.

Although no particular method can be identified as superior in every situation, the regression models generally performed well. It is worth noting that we have not performed a sensitivity analysis of the individual methods of simulating the loss triangles. By changing the inflation rate or the reporting pattern, for example, one may find that the performance of the individual methods of loss reserving will be different. We suspect, however, that the overall performance will be similar.

5 Closing Comments

Regression modeling provides an appropriate tool for estimating loss reserves. Regression methods do not provide the best answers in all situations, but are stable and have the added advantage of providing directly the variance or the confidence interval for the reserve estimate. The regression models studied are a priori fixed. In actual practice, the structure of the models will be determined from a much wider set of possible models based on an analysis of the data under review. Testing and selection of an appropriate loss reserving regression model should improve the ultimate loss reserve forecast in actual application.

Actuaries do not apply the more traditional LDF method blindly. The array of development factors is typically examined carefully before a selection of particular factors entering the reserve estimation is made. The appropriateness of the LDF method is determined for the given data set before the results of any such analysis are accepted. Professional judgment and the selection of an appropriate model are more important when regression loss reserve estimation methods are used. Therefore, an important step is missing for the regression methods as applied in this study. For the Bühlmann method, we assume knowledge of the inflation rate in addition to what is assumed known for other methods. In practice, inflation will not be known precisely, and the loss triangle will be distorted by exposure changes and inflation. This method may therefore not be as well-behaved in practice as in the simulation studies presented here.

The point estimation of the loss reserve has been the primary focus of this study, and we have not considered the variability of loss reserves around the point estimate. Verrall (1994) has outlined the procedures for computing the variance of the forecast including both the forecasting error and the parameter uncertainty.

The overall performance of the LDF method is satisfactory. The closeness of the answers of the various methods assures us that the actuarial methods of loss reserve estimation are generally well behaved. These results also tell us that regression modeling provides estimates similar to traditional actuarial methods, and one should not hesitate to use them. Given the advantage that regression methods also estimate the variability of the estimated reserve, it is expected that their use in the actuarial field will increase.

References

Bühlmann, H., Straub, E. and Schnieper, R. "Claims Reserves in Casualty Insurance Based on a Probabilistic Model." *Mitteilungen der Vereinidung Schweizerischer Versicherungsmathematiker* 1 (1980): 21–46.

Holmberg, R.D. "Correlation and the Measurement of Loss Reserve Variability." *Casualty Actuarial Society Forum* (Spring 1994): 247–277.

Mack, T. "Measuring the Variability of Chain Ladder Reserve Estimates." *Casualty Actuarial Society Forum* (Spring 1994): 101–182.

Pentikäinen, T., and Rantala, J. "A Simulation Procedure for Comparing Different Claims Reserving Methods." *Casualty Actuarial Society Forum* (Fall 1995): 128–156.

Stanard, J.N. "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques." *Proceedings of the Casualty Actuarial Society* 72 (1985): 124–148.

Verrall, R.J. "Statistical Methods for Chain Ladder Techniques." *Casualty Actuarial Society Forum* (Spring 1994): 393–446.

Zehnwirth, B. "Probabilistic Loss Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital." *Casualty Actuarial Society Forum* (Spring 1994): 447–605.